DESIGN AND DEVELOPMENT OF INTELLIGENT HYBRID RECOMMENDER SYSTEM FOR EFFECTIVE TEXT ANALYTICS

Synopsis
Submitted in the partial fulfilment for the requirements of
Doctor of Philosophy
In
Computer Science

Submitted by

PRIYA ASTHANA

Under the supervision of

Prof. Preetvanti Singh
Department of Physics and Computer Science
Faculty of science

Department of Physics and Computer Science
Faculty of Science
Dayalbagh Educational Institute (Deemed to be University)
Dayalbagh, Agra-282005

2021

Introduction

In the present fast changing life, human analysts are faced with the problem of discovering meaningful and useful information from the generated titanic of data. The increase in size is due to electronically storing the data by institutions, organizations, and business industries. Enormous text is flowing over the internet in the form of digital libraries, repositories, and social media network. With the growth of this textual data, determining appropriate patterns and trends becomes a challenging task. Thus, comes the role of text analytics, the technology of unstructured world.

1.1 Text Analytics

Text analytics scrutinizes large quantities of text, identifies critical terminology and concepts, and puts these textual data into readily searchable form. It generates high-quality information from unstructured text (patterns and trends, some combination of relevance, or interestingness) that can drive strategic decision making. Text analytics programs identify parts of sentences, determine the major themes in sentences, paragraphs, and entire documents and also determine the relationships between these. More on text analytics can be found in Fisher, Garnsey, Goel & Tam (2010); Hu & Liu (2012); Basole, Seuss & Rouse (2013); Itto & van den Bosch (2016); and Moreno & Redondo (2016). Popular software for text analytics are: Lexalytics (of Amherst, Massachusetts) that analyzes blogs; and Clarabridge (of Reston, Virginia) which analyzes consumer surveys, customer emails, and notes made on calls to service centers. Text analytics helps the analysts in

- Continuous feedback monitoring
- Identifying structural problems in products or services,
- Identifying knowledge that can avoid *reinventing the wheel*,
- Uncovering patterns and themes to reveals users' needs,
- Turning the unstructured thoughts of customers into structured data, and
- Finding hidden customer insights.

The Text Analytics Framework

Text analytics investigates unstructured contents and extracts meaningful metadata, before the content is indexed, resulting in more powerful decisions. The text analytics framework shown in Figure 1;

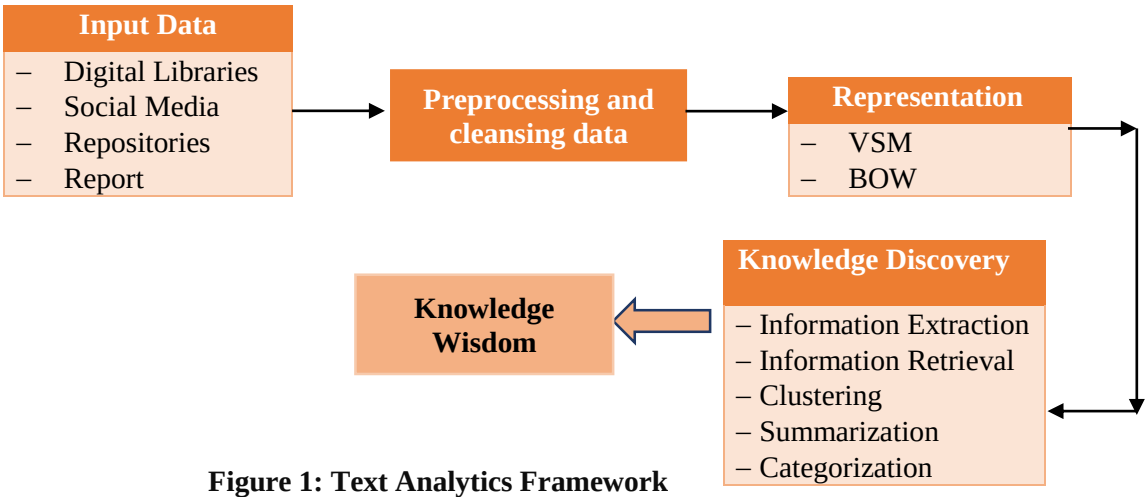

Figure 1: Text Analytics Framework

1. **Text Pre-processing:** Text pre-processing, an essential task, makes the input documents more consistent to facilitate text representation by including various text analytics techniques. These techniques include:
   - **Word sense disambiguation:** This is the ability to determine the meaning of a word with multiple definitions, often from context; for example the word *well*.
   - **Part-of-speech tagging:** This is labeling each word with a part of speech, such as noun, verb, adjective and adverb.
   - **Word stemming:** This produces the semantic root of a word by applying heuristics, for example *flying* becomes *fly*. A word stemmer looks for both forms in text.
   - **Lemmatization:** Similar to word stemming, this technique uses a more sophisticated and rigorous procedure to find word stems with more accuracy.
   - **Name entity extraction:** This locates and classifies named entities into pre-defined categories such as people, places, organizations, and so on.
   - **Relationship extraction:** This identifies how things are related.
2. **Text Representation:** Pre-processed data is then represented using *Bag Of Words* (BOW) or Vector Space Model (VSM) methods. Text representation means transforming documents into sparse numeric vectors and performing linear algebraic operations. In these basic text representation models, the linguistic structure within the text is ignored which leads to *structural curse*.
3. **Knowledge Discovery:** After successfully transforming the text corpus into numeric vectors, text mining methods like classification or clustering can be applied for knowledge discovery.

Text analytics, thus, gives business leaders the ability to extract customer insights from data sources that can provide potential for significant business impact and customer experience management (see Figure 2) that were previously too costly to analyze.

1.2 Text mining

Text mining is a process to extract interesting and significant patterns for exploring knowledge from textual data sources. It is a multi-disciplinary field based on machine learning, statistics, visualization techniques and computational linguistics. These techniques are utilized to analyze quantitative and qualitative data for predicting future performance of an organization to make effective data-driven decisions.

Text Mining Techniques

The text mining framework is illustrated in Figure 3. Various text mining techniques are:

a. Information Extraction

Information Extraction means extracting meaningful information from large amount of text. Methods are used to extract (domain) specific attributes and entities from the document and establish their relationship. The extracted corpus is stored into database for further processing.

b. Information Retrieval

Information Retrieval is a process of extracting relevant and associated patterns according to a given set of words or phrases. Different methods like query based algorithms, or probabilistic models are used to track user's behavior and search relevant data accordingly

c. Clustering

Clustering is an unsupervised process to classify the text documents in groups by applying different clustering methods like centroid, and k-means. In a cluster, similar terms or patterns are grouped extracted from various documents.

d. Text Summarization

Text summarization is a process of producing concise representation of original text documents by performing pre-processing and processing operations. Various natural language processing techniques are used for summarization.

e. Categorization

Categorization is a supervised learning method that automatically assigns predefined classes to the text documents based on their content. Machine learning tools can be used for categorization of text.

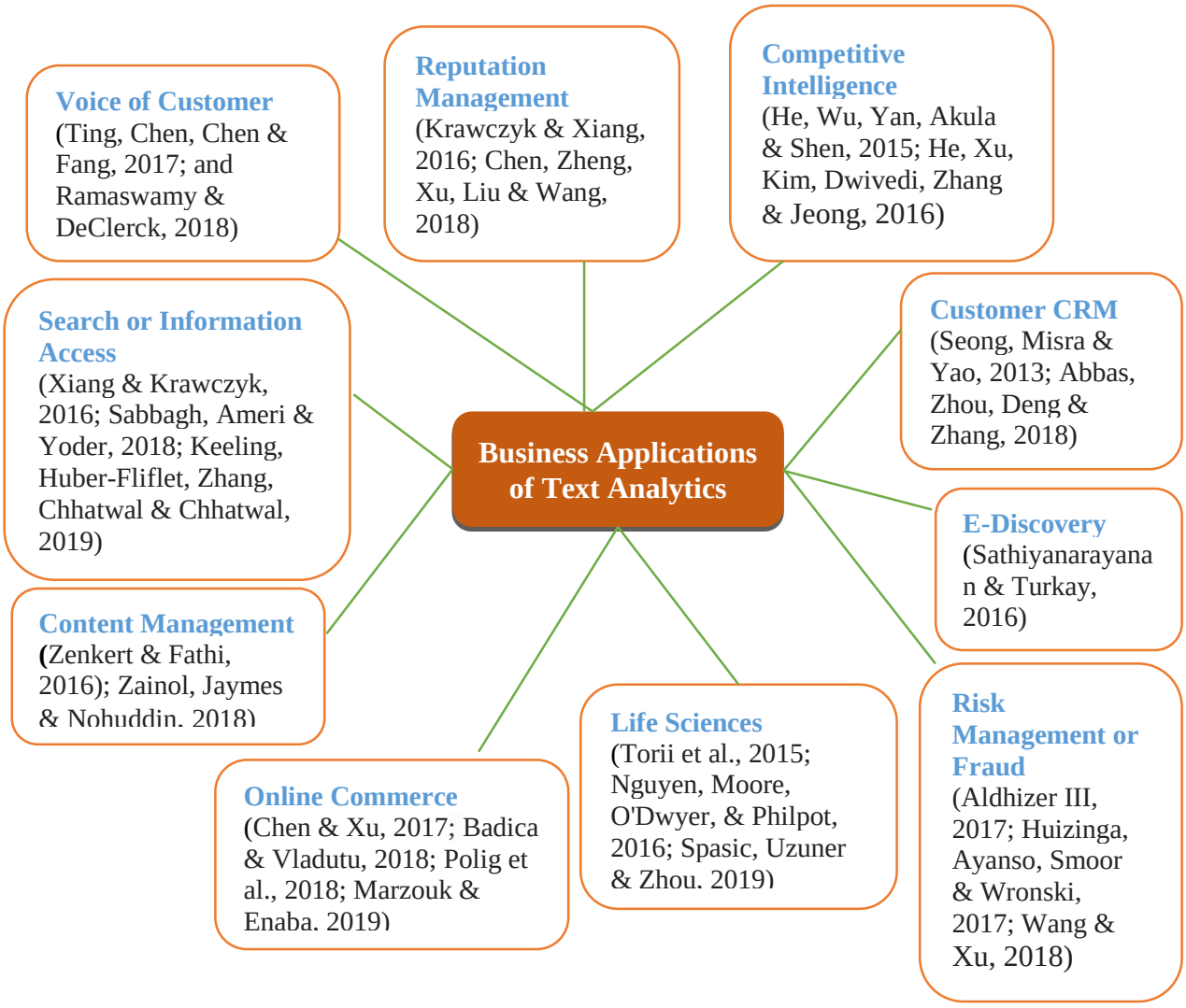
Business Applications of Text Analytics
Voice of Customer (Ting, Chen, Chen & Fang, 2017; and Ramaswamy & DeClerck, 2018)
Reputation Management (Krawczyk & Xiang, 2016; Chen, Zheng, Xu, Liu & Wang, 2018)
Competitive Intelligence (He, Wu, Yan, Akula & Shen, 2015; He, Xu, Kim, Dwivedi, Zhang & Jeong, 2016)
Search or Information Access (Xiang & Krawczyk, 2016; Sabbagh, Ameri & Yoder, 2018; Keeling, Huber-Fliflet, Zhang, Chhatwal & Chhatwal, 2019)
Customer CRM (Seong, Misra & Yao, 2013; Abbas, Zhou, Deng & Zhang, 2018)
E-Discovery (Sathiyanarayanan & Turkay, 2016)
Content Management (Zenkert & Fathi, 2016); Zainol, Jaymes & Nohuddin, 2018)
Risk Management or Fraud (Aldhizer III, 2017; Huizinga, Ayanso, Smoor & Wronski, 2017; Wang & Xu, 2018)
Life Sciences (Torii et al., 2015; Nguyen, Moore, O'Dwyer, & Philpot, 2016; Spasic, Uzuner & Zhou, 2019)
Online Commerce (Chen & Xu, 2017; Badica & Vladutu, 2018; Polig et al., 2018; Marzouk & Enaba, 2019)

Figure 2: Business Applications of Text Mining

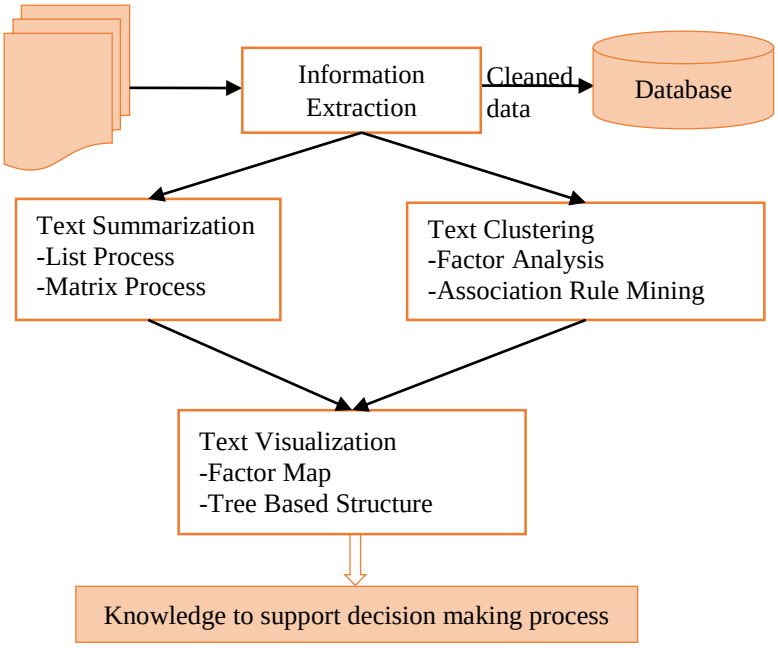

Figure 3: Text mining framework

Challenges in Text Mining

A number of issues occur during the text mining process that effect the efficiency and effectiveness of decision making. Some of these are discussed below:

a. **Integration of domain knowledge**

It is required to integrate domain knowledge (from which document corpus is to be extracted) with the computing abilities to attain the desired outcome. For this, experts need to work collaboratively from diverse domains to extract more effective, precise and accurate results. In such cases it becomes difficult to categorize the documents, as the collection of documents is large and generated from diverse fields having the same domain.

b. **Varying concepts of granularity**

This changes the context of text according to the condition and domain knowledge. There is a need to describe rules according to the domain that will be used as a standard in the area which can be embedded in text mining tools as a plug-in. This requires lots of effort and time to develop and deploy plug-ins in all fields separately, develop plug-ins in depth and have good knowledge about the domain.

c. Natural language ambiguity

Ambiguity in natural languages has lots of complications that create problem in text refinement methods and the identification of entity relationship.

Text mining is frequently used in various fields such as:

- *Digital Libraries* (Agosti, Fabris, & Silvello, 2019; Ferilli, & Angelastro, 2019) to ascertain the patterns and trends from large amounts of repositories, like journals and proceedings.
- *Academic and Research* (Comeau, Wei, Islamaj Doğan, & Lu, 2019; Lee, Yoon, Kim, Kim, Kim, So, & Kang, 2020) to analyze the educational trends, student's interest and employment ratio;
- *Life Science* (Przybyła et al. 2016; Beltagy, Cohan, & Lo, 2019) to filter out an appropriate and relevant text from a large biological repository for efficient decision making;
- *Social Media* (Bollegala, Maskell, Sloane, Hajne, & Pirmohamed, 2018; Shen, Chen & Wang, 2019) to analyze number of posts, likes and followers on the social media network for evaluating the behavior of people;
- *Business Intelligence* (Kuhzady, & Ghasemi, 2019; Saura, & Bennett, 2019) to analyze business information of organizations and enterprises.

1.3 Recommender System

Recommender systems, growing popular in recent years, are computerized systems utilized to provide suggestions associated with various decision-making processes to users according to their requirements. It is defined as a decision-making strategy for users under complex information environments. The recommendation systems are used by Wu, Wall, & Pearce (2014); Han, Mankad, Gavirneni & Verma (2016); and Xia, Goldberg, Hong & Garvey (2019) to solve various real-world problems.

Recommender systems are information filtering systems that filter vital information fragment out of dynamically generated information according to user's preferences, interests, or observed behavior about an item. Recommender systems function with two kinds of information:

- *Characteristic information*: This is the information about items (like keywords, categories) and users (like preferences, profiles).

- *User-item interactions*: This includes information such as ratings, number of purchases, likes, and so on.

Recommendation techniques can be divided into following five categories:

- **Collaborative-filtering:** In this approach, recommendation for each active user is received by comparing with the preferences of other users who have rated the product in a similar way as the active user. The basic assumption here is that people who had similar tastes in the past will also have similar tastes in the future. Ratings or user generated feedbacks are used to determine taste commonalities between groups of users and generate recommendations based on inter-user similarities. This type of filtering suffers from problems like cold-start (new user or new item), and *gray sheep* (users that do not fit in any taste cluster).
- **Content-Based filtering:** Here recommendations depend on users' former choices. Item description and profile of the user's orientation plays an important role in this type of filtering. Recommendation quality is limited by the selected features of the recommended items. This method too suffers from the cold-start problem.
- **Demographic Filtering:** Demographic filtering recommendations are based on a demographic profile of the user. Here the challenge is collecting enough demographic information that is needed because of online privacy concerns.
- **Knowledge-Based Filtering:** This type of filtering generates recommendations by using knowledge about users and items that meet the users' requirements.
- **Hybrid filtering:** The hybrid filtering method combines the advantages of above methods, thus reducing problems that are associated with individual filtering approaches such as cold start problem, overspecialization problem and sparsity problem. Implementation of hybrid filtering method improves the accuracy and efficiency of recommendation process.

  On the basis of these filtering approaches, Recommender Systems can be categorized as (Figure 4):

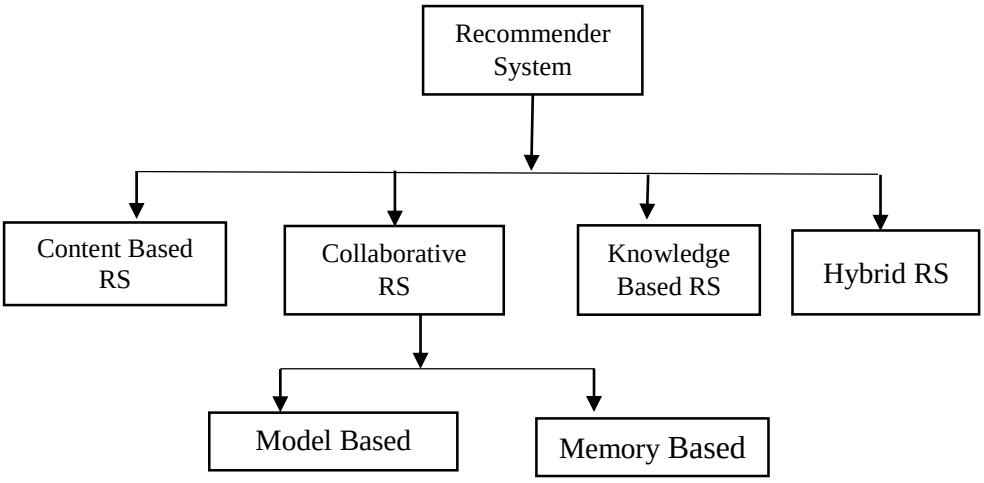

Figure 4: Types of Recommender System (RS)

a. Content-based recommender systems

This system uses characteristic information. Users or items have profiles describing their characteristics and the system recommends an item to a user if the two profiles match. User profiles are constructed using historical interactions or by explicitly asking users about their interests. The main disadvantage of these systems is their strong dependence on an application domain.

b. Collaborative based recommender systems

These systems are based on user-item interactions. The system uses historic users' preference for items (for example watched, purchased, liked), which can be presented as a user-item matrix where the rows present the users and the items are presented column wise. Collaborative filtering algorithms are divided into two sub-categories:

- **Model Based Method**

  Model based collaborative approaches rely only on user-item interactions information and assume a latent model supposed to explain these interactions.

- **Memory Based Method**

  Memory based methods directly work with values of recorded interactions, assuming no model, and are essentially based on nearest neighbour search.

c. Knowledge based recommender system

In knowledge-based systems, recommendations are based on the conclusions about the needs and preferences of the user. Usually, this knowledge is manually added by experts, but it can also contain functional dependencies.

d. Hybrid recommender system

Hybrid recommender systems combine various recommendation strategies in different ways. Hybrid method gives the best accuracy as compared to collaborative and content-based method.

Development of hybrid recommender system using various text mining techniques enables decision maker and analysts to make effective decisions.

Review of Literature

In recent years, text analytics is gaining popularity because the analysis of the text content in forms of textual communication helps in analyzing customers comments, or perceptions of a company, brand, or product. Das (2014) surveyed the technology and empirics of text analytics in finance and presented tools for information extraction.

2.1 Text Analytics

Endert, Fiaux & North (2012) designed visual analytic interaction system to enable analyst for spatial interaction. Seong, Misra & Yao (2013) performed customer segmentation analysis in combination with Recency, Frequency and Monetary, a method used for determining the customer lifetime value. Zheng, Zhu & Lin (2013) developed a semi-supervised system for estimating quality of online reviews. Brewster, Polovina, Rankin & Andrews (2014) utilized open–source data analytics and ontological knowledge representation law enforcement agencies to deal with human trafficking and other organized crimes. Khalili, Auer & Ngomo (2014) developed a text analytics architecture of participation to enable users use sophisticated techniques for analyzing and visualizing their content. Poulin et al. (2014) developed linguistics-driven prediction models to estimate the risk of suicide from unstructured clinical notes. Wu, Wall, & Pearce (2014) examined experiences of international tourists through Leximancer. He, Wu, Yan, Akula & Shen (2015) proposed a social media competitive analytics framework with sentiment benchmarks. Markham, Kowolenko & Michaelis (2015) described methods to use unstructured text analytics for supporting critical product development decisions. Nitin, Swapna &

Shankararaman (2015) developed student feedback mining system by using text analytics and opinion mining approaches to provide instructors a quantified and exhaustive analysis of the qualitative feedback from students. Tussyadiah & Zach (2015) explained the key service attributes sought by guests by using terminology extraction and word co-occurrence networks to support the competitive advantage of hotels and P2P rentals. Xiang, Schwartz, Gerdes Jr & Uysal (2015) applied a text analytical approach to examine reviews extracted from Expedia.com. Xiang, Schwartz & Uysal (2015) utilized text analytics approaches for analyzing customer reviews to understand the preferences of hotels by the guest.

He et al. (2016) analyzed public social media data to mine competitive intelligence in social media for determining the market insight by comparing consumer opinions and sales performance of a business. Müller, Junglas, Debortoli & vom Brocke (2016) applied text analytics to the contents of service requests in three organizations, and concluded that inspite of each organization having a different analytical goal, their experience was similar. Nguyen, Moore, O'Dwyer & Philpot (2016) assessed the utility of Medtex on automating Cancer Registry notifications from narrative histology and cytology. Sathiyanarayanan & Turkay (2016) proposed a multi-modal and multi-level approach for generating automated visual representations using a keyword search facility to extract relevant information from the email. Yuan, Lau & Xu (2016) designed text analytics-based framework to develop the Domain-Constraint Latent Dirichlet Allocation topic model for effective extraction of topical features from texts. Zenkert & Fathi (2016) explained the method to integrate text analytics and knowledge base. Aldhizer III (2017) detailed the benefits of using visual and text analytics to explore and visualize structured data for material misstatements and fraud. Chen & Xu (2017) developed a method to extract data by combining domain ontology and topic text analytics approach. Huizinga, Ayanso, Smoor & Wronski (2017) explored twitter data of insurance and natural disasters to gain business insights using text analytics. Reddick, Chatfield & Ojo (2017) developed a framework for facilitating organizational learning through social media text analytics to enhance citizen-centric public service quality. Zainol, Nohuddin, Mohd & Zakaria (2017) developed a method of text analytics on peacekeeping documents to discover significant text patterns.

Abbas, Zhou Deng & Zhang (2018) proposed a LAP-based text analytics framework to support sense-making in online discourse. Badica & Vladutu (2018) developed a sentiment analysis algorithm for

examining the relations between the online review rating score and helpfulness. Chen, Zheng, Xu, Liu & Wang (2018) applied text analytics methods to analyze the reputation of products, online markets and hotel brands. Das, Sahoo, Pandey (2018) proposed CV parser model using text analytics to extract entities needed for recruitment process in companies. Polig et al. (2018) presented a framework to transform text analytics queries into a hardware description. Ramaswamy & DeClerck (2018) explored different technologies of text analytics to analyze the contextual information for capturing customer feedback efficiently. Sabbagh, Ameri, & Yoder (2018) proposed a novel framework for capability-based supplier classification that relies on the unstructured capability narratives available on the suppliers websites. Khedkar, Shinde (2018) proposed text analytics based Ensemble Classifier to classify review sentences as praise or complaint. Wang & Xu (2018) proposed a novel deep learning model for automobile insurance fraud detection using Latent Dirichlet Allocation based text analytics. Zainol, Jaymes & Nohuddin (2018) developed a text analytics tool, VisualUrText, to assist students and researchers in document analyses. Zhou, Qiao, Du, Wang, Fan & Yan (2018) developed a singular value decomposition-based semantic keyword similarity method to quantify customer agility using large-scale customer review texts and product release notes.

Ibrahim, & Wang (2019) used topic modelling, sentiment analysis, and network analysis to analyze the tweets for identifying the most-discussed topics. Keeling, Huber-Fliflet, Zhang, Chhatwal & Chhatwal (2019) performed text analytics on data sets and search term lists from real legal matters to determine effective search terms. Marzouk & Enaba (2019) developed a dynamic text analytical model for contract and correspondence to assist project parties in exploring the project performance. Transfer learning has been widely utilized in many machine learning applications. Xia, Goldberg, Hong & Garvey (2019) incorporated three different types of knowledge transfer approaches into a deep learning-based healthcare text analytics model.

2.2 Text Mining

Text mining is an important process for text analytics. He (2013) applied data mining and text mining to analyze the online questions and chat messages automatically recorded by a live video streaming system. He, Zha & Li (2013) utilized text mining to analyze unstructured text content on Facebook and Twitter sites of three largest pizza chains. Mostafa (2013) applied text mining techniques on tweets to

evaluate consumers' sentiment towards well-known mobile brands. Wei, Kao & Lu (2013) developed Pub Tator, a web-based system for assisting biocuration based on multiple challenge-winning text mining algorithms. Nassirtoussi, Aghabozorgi, Wah & Ngo (2014) reviewed market prediction based online-text-mining procedures. Suarez-Tangil, Tapiador, Peris-Lopez & Blasco (2014) developed a system using information retrieval technique for classifying code structures in android malware families. Jha (2018) developed a method to recognize emotion from sms and social media text. Chaix, Deléger, Bossy & Nédellec (2019) described a procedure for applying text mining tools to a large collection of PubMed abstracts in the food microbiology domain. Gunay, Shen & Yang (2019) developed a text-mining method to extract information about failure patterns in building systems and components from various databases. Hong & Park (2019) performed text mining analysis to extract major keywords, and confirmed differences in influences of the keywords which affect corporate performance. Kim & Chung (2019) proposed a method for extracting associative feature information using text mining from health big data. Lamurias & Couto (2019) presented an overview of biomedical text mining tools. Liu (2019) proposed a Chinese text classification model for the project management office using fuzzy semantics and text mining techniques. Loureiro, Guerreiro, Eloy, Langaro & Panchapakesan (2019) reviewed text mining for the use of virtual reality in marketing. Ramanathan & Meyyappan (2019) developed a sentiment analysis method based on common sense knowledge for Oman tourism. Saura & Bennett (2019) proposed a three-step research methodology based on data text to analyze user generated content in social networks. Sezgen, Mason & Mayer (2019) investigated the key drivers of customer satisfaction. Text mining and natural language process techniques were applied by Zhang, Fleyeh, Wang & Lu (2019) to analyze the construction accident reports. Zhou, Ng, Lee, Xu & Yang (2019) proposed a text mining approach to minimize information asymmetry between technology developers and AEC users. Lee et al. (2020) introduced pre-trained language model that can be adapted for biomedical corpora. A Text Mining Analysis Roadmap was developed by Zaki & McColl-Kennedy (2020) to provide efficient recommendations.

Recommender System

A recommender system is an information filtering system. Hybrid recommender system takes advantages over standalone recommender system in generating effective recommendations. Some

reviews on hybrid recommender System are presented by Cano & Morisio(2017); Aerts, Claes, Leeuw & Leeuw(2019); & Shokeen & Rana(2020). Lin, Lai, Chen & Hwang (2015) utilized text-mining techniques to construct user interest profiles for making personalized hotel recommendations. Maatallah & Seridi-Bouchelaghem (2015) proposed a fuzzy hybrid diversified recommendation system able to generate multi-taste recommendations depending on the user's profile variation. A hierarchical Bayesian model was developed by Wang, Wang & Yeung (2015) to perform deep representation learning for the content information and collaborative filtering. Capdevila, Arias & Arratia (2016) presented a hybrid recommender system for a popular location-based social network where users are able to write short reviews on the places of interest they visit. Farokhi, Vahid, Nilashi & Ibrahim (2016) aimed to improve the recommendation accuracy of Collaborative Filtering techniques for tourism recommender systems. Strub, Gaudel & Mary (2016) discussed the procedure to develop hybrid recommender system based on autoencoder and concluded that this approach improves the architecture.

Logesh & Subramaniyaswamy (2019) developed a Personalized Context-Aware Hybrid Travel Recommender System by incorporating user's contextual information and evaluated the system on the real-time large-scale dataset of Yelp and Trip Advisor. Logesh, Subramaniyaswamy, Vijayakumar & Li (2019) proposed activity and behavior induced Personalized Recommender System as a hybrid approach to predict persuasive POI recommendations. An emotional fingerprint Based Model for movies recommendation system was proposed by Nosshi, Asem & Senousy (2019). Qian, Zhang, Ma, Yu & Peng (2019) proposed an emotion-aware recommender system based on hybrid information fusion to comprehensively analyze the user's features. Tsai & Brusilovsky (2019) discussed methods for transparency of recommendations by augmenting an interactive hybrid recommender interface with several types of explanations. Vall et al. (2019) introduced two feature-combination hybrid recommender systems by integrating the collaborative information encoded in curated music playlists with any type of song feature vector representation. Haubner & Setzer (2020) proposed a method for learning weighting schemes in weighted hybrid recommender systems that is based on statistical forecast and portfolio theory. Web Page health recommender systems were introduced by Manikandan & Saravanan (2020) to provide appropriate web pages for patients. Oyebode & Orji (2020) proposed a hybrid recommender system that combines the item-based collaborative filtering technique and the

demographic-based approach. Pan, He, Yu, & Li (2020) proposed an adaptive method to learn the trust influence between users with multiple roles of truster and trustee for recommendation.

Motivation

From the above discussion the following observations were made:

- *While extracting meaning from text and other unstructured sources is complex, companies are realizing the value of turning digital customer chatter into actionable business intelligence. Companies that underestimate the value of unstructured data do so at their peril* (Müller, Junglas, Debortoli, & vom Brocke, 2016).

  Improving the accuracy of analyzing large volumes of text is of utmost importance to companies, institutions, and organizations, and one of the best ways to do this is incorporate text analytics into the back end.

- *The modern world generates enormous amounts of data. This data has become the most valuable managerial resource to provide a competitive edge and originate knowledge management initiatives* (Omotayo, 2015).

  Automation of this data becomes essential to derive high quality information. Text Mining is an approach which can be used for more efficient search.

- *Transforming masses of unstructured information — documents, audio, video, and images — into quantifiable data can be processed and analyzed by business decision makers* (Rusu et al., 2013; Mirończuk, 2020).

  The main difference of such transformation from other types of data analysis is that the input data is not formalized in any way, which means it cannot be described with a simple mathematical function. Thus, methods may be developed to transform this unstructured information into quantitative functions.

- *Historically people have relied on their peers or on experts' suggestions for decision support and recommendations about commodities, news or entertainment, etc. The exponential growth of the digital information has created the problem of information overload* (Çano & Morisio, 2017).

Developing a recommender system becomes necessary to overcome the problem of conjunction with the information overload.

- *Often the recommender system is unaware of the actual intent of the user and simply provides recommendations dynamically without properly understanding the thought process of the user* (Bhattacharya, Burhanuddin, Sancheti, & Satya, 2017).

  An intelligent recommender system is not only useful for the user but also for businesses which want to learn the tendencies of their users.

Proposed Work

The main goal of this research study is to develop an intelligent hybrid recommendation system for effective text analytics.

Objectives of the proposed study

- To develop method for transforming unstructured text into quantifiable functions,
- To develop text mining tool for analyzing quantifiable/structured business data,
- To design and develop intelligent recommender system for facilitating organizational learning,
- To provide a real-life implementation of the system.
- To compare the performance of proposed system with the existing system.

Methodology to achieve the objectives

The major tasks of the systems are:

- To achieve the 1st objective, text analytics tools will be used to quantify the unstructured data. This study will deal with the data based on the application area chosen for the study area.
- For attaining the 2nd objective, text mining tools such as clustering, summarization will be integrated with other decision making tools for analyzing the business data.
- The 3rd objective of the study will be accomplished by incorporating the above developed tools within a hybrid recommender system to provide intelligence to the system. The variables and parameters for DSS will be selected as per the study area.
- The developed system will be implemented for analyzing the healthcare data or academic records (taking Dayalbagh Educational Institute as a case study).

- The performance of developed system will be compared with the existing system by using evaluation metrics.

Stages of building proposed system

The development stages of the system are as follows (Figure 2):

1. **Planning and feasibility analysis**: This study will plan a strategic analysis, including the evaluation of organization business lines. A number of designing, developing and refining iterations will take place according to the tactical and strategic business requirements.
2. **Requirement analysis:** This is an important stage in the development cycle. Requirements of domain experts will be identified to acquire the basic understanding of the application domain for formulating varied requirements. Then the business subject priorities will be established, depending on their relative importance, and feasibility of the necessary data. This priority list will used to establish the subsequent iterations of the system development.
3. **Design and Analysis:** Based on the established priorities, the identified requirements will be analyzed and refined for detailed understanding. Conceptual model of the system will be developed using UML diagrams.

   Analysis artifacts will be refined so that these can be implemented. The relevant data will be collected from primary and secondary sources. Database infrastructure will be created to include all the information necessary for the implementation.

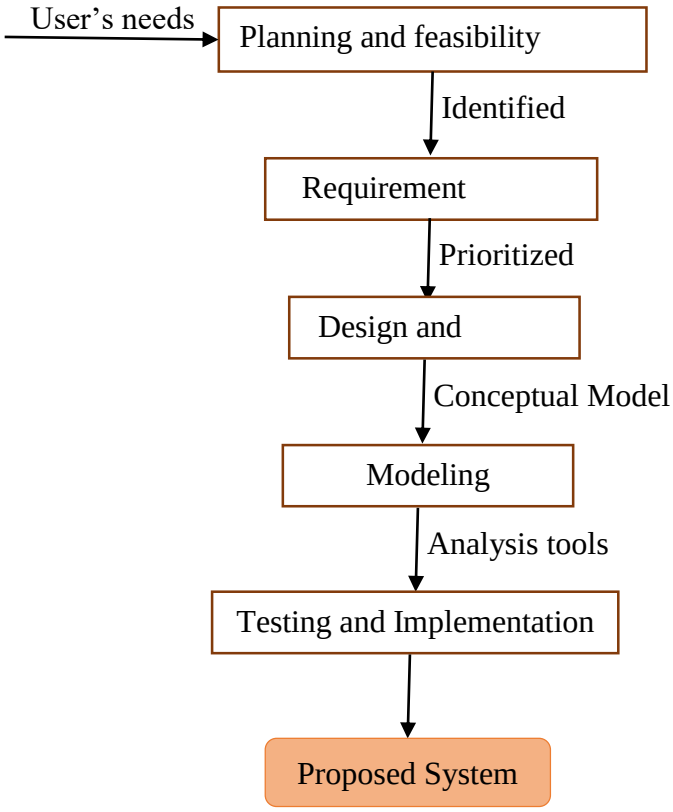

Figure 5: Stages of development

The application domain for this study will be one of the following (depending on the availability of the data):

a. Generating smart recommendations in the field of higher education (taking Dayalbagh Educational Institute (Deemed University), Agra as a case study);

b. Generating smart recommendations in the field of healthcare (taking Agra as a case study).

4. **Modeling:** Techniques like data mining, ontology, multi-criteria decision making, and/or case-based reasoning will be integrated for effective knowledge modeling to provide smart decisions. The developed system will consist of Database, knowledge warehouse, and Model base. In order to develop knowledge warehouse a semantic based approach will be used. The main idea behind this this will be to extract and use knowledge from the data. The developed knowledge warehouse will provide the decision maker with an intelligent analysis platform that will enhance all phases of the knowledge management process. Intelligent analysis platform will provide array of hybrid knowledge modeling and analysis tools (Figure 6).

5.

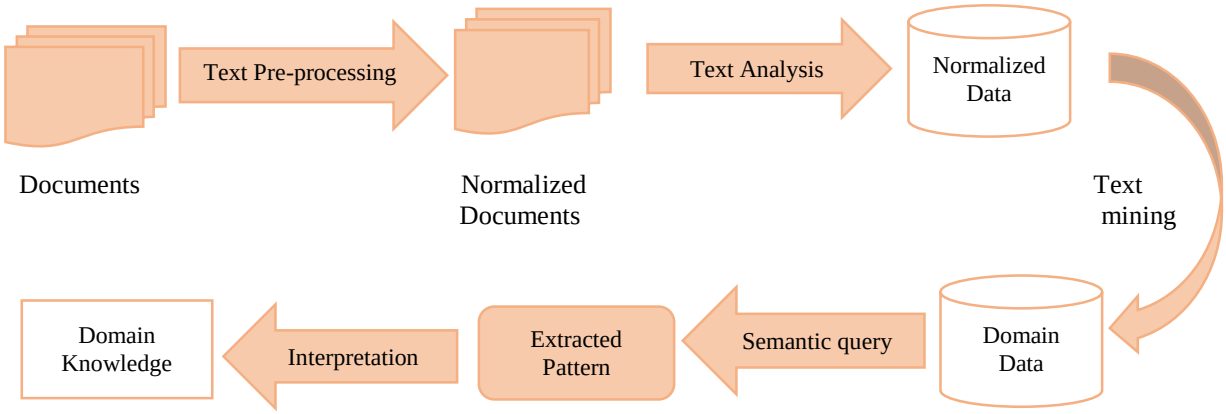

Figure 6: Text mining process

6. **Testing and implementation:** Once the planning and design stages are completed, the current iteration for system implementation will start. In this stage, the development and testing environments will be established, the hardware and software components will be installed, and the configuration management process will be implemented.

Strategies used in Hybrid Recommender System

Some of the strategies used in hybrid recommender system are as follows:

1. **Weighted**

   In weighted hybrid recommender system the recommended item's score is obtained by aggregates the results of all combined approaches. Initially system assigned equal weights to all recommenders which is adjusted progressively according to the predictions.

2. **Switching**

   In switching recommender system we can switch between the different recommendation techniques if a particular recommendation technique does not provide the accurate recommendation to the user then another method is attempted.

3. **Mixed**

   In mixed hybrid recommender system various recommendations techniques are used to make the final recommendation. In this several independent list of recommendations produced by different technique and mixed hybrid recommender generates the consolidated list.

4. **Cascade**

   In cascade recommender system various recommendation techniques are used if one technique is used to ranking the user or items and other is used to refines the recommendation. The main advantage of this method is weak technique with low priority cannot alter the decision made by strong technique of higher priority.

5. **Feature Augmentation**

   In feature augmentation technique the result produced by first technique is incorporated with the next recommendation. This technique is used when we need additional feature or information. It also allows us to combine two recommender systems in such a way that output of first recommender system is used as input to second recommender system.

Some of these strategies will be applied while developing the proposed Recommender System depending upon the chosen application area.

Evaluation metric for Recommender System

Confusion matrix or some other evaluation metric will be used to evaluate the performance of developed algorithm.

To conclude, the proposed system will be useful for various fields i.e. academics, e-commerce, and e-learning etc. The developed text analytics-based intelligent recommendation system will be helpful for decision-makers by recommending the appropriate suggestion which helps in the smart decision-making process.

References

Abbas, A., Zhou, Y., Deng, S., & Zhang, P. (2018). Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly*, *42*(2).

Aerts, L., Claes, A., Leeuw, D. D., & Leeuw, E. D. (2019, March). Implicit hybrid recommender system for informative articles. In *Seminar case studies in business analytics and quantitative marketing (FEM21001)* (pp. 1-52).

Agosti, M., Fabris, E., & Silvello, G. (2019, January). On synergies between information retrieval and digital libraries. In *Italian Research Conference on Digital Libraries* (pp. 3-17). Springer, Cham.

Aldhizer III, G. R. (2017). Visual and Text Analytics: The Next Step in Forensic Auditing and Accounting. *The CPA Journal*, *87*(6), 30-33.

Badica, C., & Vladutu, G. (2018, September). Application of Meaningful Text Analytics to Online Product Reviews. In *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (pp. 327-334). IEEE.

Basole, R. C., Seuss, C. D., & Rouse, W. B. (2013). IT innovation adoption by enterprises: Knowledge discovery through text analytics. *Decision Support Systems*, *54*(2), 1044-1054.

Beltagy, I., Cohan, A., & Lo, K. (2019). Scibert: Pretrained contextualized embeddings for scientific text. *arXiv preprint arXiv:1903.10676*.

Bhattacharya, B., Burhanuddin, I., Sancheti, A., & Satya, K. (2017, November). Intent-aware contextual recommendation system. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 1-8). IEEE.

Bollegala, D., Maskell, S., Sloane, R., Hajne, J., & Pirmohamed, M. (2018). Causality patterns for detecting adverse drug reactions from social media: text mining approach. *JMIR public health and surveillance*, *4*(2), e51.

Brewster, B., Polovina, S., Rankin, G., & Andrews, S. (2014, July). Knowledge management and human trafficking: Using conceptual knowledge representation, text analytics and open-source data to combat organized crime. In *International Conference on Conceptual Structures* (pp. 104-117). Springer, Cham.

Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, *21*(6), 1487-1524.

Capdevila, J., Arias, M., & Arratia, A. (2016). GeoSRS: A hybrid social recommender system for geolocated data. *Information Systems*, *57*, 111-128.

Chaix, E., Deléger, L., Bossy, R., & Nédellec, C. (2019). Text mining tools for extracting information about microbial biodiversity in food. *Food microbiology*, *81*, 63-75.

Chen, R., & Xu, W. (2017). The determinants of online customer ratings: a combined domain ontology and topic text analytics approach. *Electronic Commerce Research*, *17*(1), 31-50.

Chen, R., Zheng, Y., Xu,W.,Liu,M.,& Wang,J.(2018). Seconhhand seller reputation in online markets: A text analytics framework. Decision Support Systems, 108,96-10

Comeau, D. C., Wei, C. H., Islamaj Doğan, R., & Lu, Z. (2019). PMC text mining subset in BioC: about three million full-text articles and growing. *Bioinformatics*, *35*(18), 3533-3535.

Das, S. R. (2014). Text and context: Language analytics in finance. *Foundations and Trends® in Finance*, *8*(3), 145-261.

Endert, A., Fiaux, P., & North, C. (2012, May). Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 473-482).

Farokhi, N., Vahid, M., Nilashi, M., & Ibrahim, O. (2016). A multi-criteria recommender system for tourism using fuzzy approach. *Journal of Soft Computing and Decision Support Systems*, *3*(4), 19-29.

Ferilli, S., & Angelastro, S. (2019, January). Towards a process mining approach to grammar induction for digital libraries. In *Italian Research Conference on Digital Libraries* (pp. 291-303). Springer, Cham.

Fisher, I. E., Garnsey, M. R., Goel, S., & Tam, K. (2010). The role of text analytics and information retrieval in the accounting domain. *Journal of emerging technologies in Accounting*, *7*(1), 1-24.

Gunay, H. B., Shen, W., & Yang, C. (2019). Text-mining building maintenance work orders for component fault frequency. *Building Research & Information*, *47*(5), 518-533.

Han, H. J., Mankad, S., Gavirneni, N., & Verma, R. (2016). What guests really think of your hotel: Text analytics of online customer reviews.

Haubner, N., & Setzer, T. (2020, January). Applying Optimal Weight Combination in Hybrid Recommender Systems. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, *52*(7), 801-812.

He, W., Xu, G., Kim, Y., Dwivedi, R., Zhang, J., & Jeong, S. R. (2016). Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online Information Review*.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, *29*(1), 90-102.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464-472.

Hong, J. W., & Park, S. B. (2019). The Identification of Marketing Performance Using Text Mining of Airline Review Data. *Mobile Information Systems*, *2019*.

Hu, X., & Liu, H. (2012). Text analytics in social media. In *Mining text data* (pp. 385-414). Springer, Boston, MA.

Huizinga, T., Ayanso, A., Smoor, M., & Wronski, T. (2017). Exploring insurance and natural disaster tweets using text analytics. *International Journal of Business Analytics (IJBAN)*, *4*(1), 1-17.

Ibrahim, N. F., & Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems*, *121*, 37-50.

Ittoo, A., & van den Bosch, A. (2016). Text analytics in industry: Challenges, desiderata and trends. *Computers in Industry*, *78*, 96-107.

Keeling, R., Huber-Fliflet, N., Zhang, J., Chhatwal, R. P., & Chhatwal, P. (2019). Separating the Privileged Wheat from the Chaff–Using Text Analytics and Machine Learning to Protect Attorney-Client Privilege. *Richmond Journal of Law & Technology*, *25*(3).

Khalili, A., Auer, S., & Ngomo, A. C. N. (2014, May). context–lightweight text analytics using linked data. In *European Semantic Web Conference* (pp. 628-643). Springer, Cham.

Kim, J. C., & Chung, K. (2019). Associative feature information extraction using text mining from health big data. *Wireless Personal Communications*, *105*(2), 691-707.

Krawczyk, M., & Xiang, Z. (2016). Perceptual mapping of hotel brands using online reviews: a text analytics approach. *Information Technology & Tourism*, *16*(1), 23-43.

Kuhzady, S., & Ghasemi, V. (2019). Factors influencing customers' satisfaction and dissatisfaction with hotels: A text-mining approach. *Tourism Analysis*, *24*(1), 69-79.

Lamurias, A., & Couto, F. (2019). Text mining for bioinformatics using biomedical literature. *Encyclopedia of bioinformatics and computational biology*, *1*, 602-611.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, *36*(4), 1234-1240.

Lin, K. P., Lai, C. Y., Chen, P. C., & Hwang, S. Y. (2015, October). Personalized hotel recommendation using text mining and mobile browsing tracking. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 191-196). IEEE.

Liu, J. W. (2019). Using big data database to construct new GFuzzy text mining and decision algorithm for targeting and classifying customers. *Computers & Industrial Engineering*, *128*, 1088-1095.

Logesh, R., & Subramaniyaswamy, V. (2019). Exploring hybrid recommender systems for personalized travel applications. In *Cognitive informatics and soft computing* (pp. 535-544). Springer, Singapore.

Logesh, R., Subramaniyaswamy, V., Vijayakumar,V., & Li, X.(2019). Effiicent user profiling based intelligent travel recommender system for individual and group of users. Mobile Networks and Applications, 24(3),1018-1033.

Loureiro, S. M. C., Guerreiro, J., Eloy, S., Langaro, D., & Panchapakesan, P. (2019). Understanding the use of Virtual Reality in Marketing: A text mining-based review. *Journal of Business Research*, *100*, 514-530.

Manikandan, R., & Saravanan, V. (2020). A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective. *Multimedia Tools and Applications*, *79*(5), 3807-3829.

Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management*, *58*(2), 30-39.

Marzouk, M., & Enaba, M. (2019). Text analytics to analyze and monitor construction project contract and correspondence. *Automation in Construction*, *98*, 265-274.

Maatallah, M., & Seridi-Bouchelaghem, H. (2015). A fuzzy hybrid approach to enhance diversity in top-N recommendations. *International Journal of Business Information Systems*, *19*(4), 505-530.

Mirończuk, M. M. (2020). Information Extraction System for Transforming Unstructured Text Data in Fire Reports into Structured Forms: A Polish Case Study. *Fire Technology*, *56*(2), 545-581.

Moreno, A., & Redondo, T. (2016). Text analytics: the convergence of big data and artificial intelligence. *IJIMAI*, *3*(6), 57-64.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, *40*(10), 4241-4251.

Müller, O., Junglas, I., Debortoli, S., & vom Brocke, J. (2016). Using text analytics to derive customer service management benefits from unstructured data. *MIS Quarterly Executive*, *15*(4), 243-258.

N. Jha, "An Approach towards Text to Emoticon Conversion and Vice-Versa Using NLTK and WordNet," in *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*, Changsha, China, 2018 pp. 161-166.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, *41*(16), 7653-7670.

Nguyen, A. N., Moore, J., O'Dwyer, J., & Philpot, S. (2016). Automated cancer registry notifications: validation of a medical text analytics system for identifying patients with cancer from a state-wide pathology repository. In *AMIA Annual Symposium Proceedings* (Vol. 2016, p. 964). American Medical Informatics Association.

Nitin, G. I., Swapna, G., & Shankararaman, V. (2015, October). Analyzing educational comments for topics and sentiments: A text analytics approach. In *2015 IEEE frontiers in education conference (FIE)* (pp. 1-9). IEEE.

Nosshi, A., Asem, A. S., & Senousy, M. B. (2019). Hybrid Recommender System Using Emotional Fingerprints Model. *International Journal of Information Retrieval Research (IJIRR)*, *9*(3), 48-70.

Omotayo, F. O. (2015). Knowledge Management as an important tool in Organisational Management: A Review of Literature. *Library Philosophy and Practice*, *1*(2015), 1-23.

Oyebode, O., & Orji, R. (2020). A hybrid recommender system for product sales in a banking environment. *Journal of Banking and Financial Technology*, 1-11.

Pan, Y., He, F., Yu, H., & Li, H. (2020). Learning adaptive trust strength with user roles of truster and trustee for trust-aware recommender systems. *Applied Intelligence*, *50*(2), 314-327.

P. Das, B. Sahoo and M. Pandey, "A Review on Text Analytics Process with a CV Parser Model," *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, 2018, pp. 1-7.

Polig, R., Atasu, K.,& Giefers, H., & Hagleitner, C., Chiticariu, L., Reiss, F., & Hofstee, P. (2018). A hardware compilation framework for text analytics queries. Journal of Parallel and Distributed Computing. 111, 260-272

Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goertzel, B., ... & McAllister, T. (2014). Predicting the risk of suicide by analyzing the text of clinical notes. *PloS one*, *9*(1).

Przybyła, P., Shardlow, M., Aubin, S., Bossy, R., Eckart de Castilho, R., Piperidis, S., ... & Ananiadou, S. (2016). Text mining resources for the life sciences. *Database*, *2016*.

Qian, Y., Zhang, Y., Ma, X., Yu, H., & Peng, L. (2019). EARS: Emotion-aware recommender system based on hybrid information fusion. *Information Fusion*, *46*, 141-146.

Ramanathan, V., & Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about Oman tourism. In *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1-5). IEEE.

Ramaswamy, S., & DeClerck, N. (2018). Customer Perception Analysis Using Deep Learning and NLP. *Procedia Computer Science*, *140*, 170-178.

Reddick, C. G., Chatfield, A. T., & Ojo, A. (2017). A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government Facebook use. *Government Information Quarterly*, *34*(1), 110-125.

Rusu, O., Halcu, I., Grigoriu, O., Neculoiu, G., Sandulescu, V., Marinescu, M., & Marinescu, V. (2013, January). Converting unstructured and semi-structured data into knowledge. In *2013 11th RoEduNet International Conference* (pp. 1-4). IEEE.

S.Atul Khedkar and S. K. Shinde, "Customer Review Analytics for Business Intelligence," *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Madurai, India, 2018, pp. 1-5.

Sabbagh, R., Ameri, F., & Yoder, R. (2018). Thesaurus-guided text analytics technique for capability-based classification of manufacturing suppliers. *Journal of Computing and Information Science in Engineering*, *18*(3).

Sathiyanarayanan, M., & Turkay, C. (2016). Determining and Visualising E-mail Subsets to Support E-discovery.

Saura, J. R., & Bennett, D. R. (2019). A Three-Stage method for Data Text Mining: Using UGC in Business Intelligence Analysis. *Symmetry*, *11*(4), 519.

Seong, K. T., Misra, A. H., & Yao, J. J. (2013). Be Customer Wise or Otherwise: Combining data mining and interactive visual analytics to analyze large and complex customer resource management (CRM) data. In *SAS Global Forum* (pp. 1-11).

Sezgen, E., Mason, K. J., & Mayer, R. (2019). Voice of airline passenger: A text mining approach to understand customer satisfaction. *Journal of Air Transport Management*, *77*, 65-74.

Shen, C. W., Chen, M., & Wang, C. C. (2019). Analyzing the trend of O2O commerce by bilingual text mining on social media. *Computer*

s in Human Behavior, *101*, 474-483.

Shokeen, J., & Rana, C. (2020). A study on features of social recommender systems. *Artificial Intelligence Review*, *53*(2), 965-988.

Spasic, I., Uzuner, Ö., & Zhou, L. (2019). Emerging clinical applications of text analytics. *International journal of medical informatics*.

Strub, F., Gaudel, R., & Mary, J. (2016, September). Hybrid recommender system based on autoencoders. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems* (pp. 11-16).

Suarez-Tangil, G., Tapiador, J. E., Peris-Lopez, P., & Blasco, J. (2014). Dendroid: A text mining approach to analyzing and classifying code structures in android malware families. *Expert Systems with Applications*, *41*(4), 1104-1117.

Ting, P. J. L., Chen, S. L., Chen, H., & Fang, W. C. (2017). Using big data and text analytics to understand how customer experiences posted on yelp. com impact the hospitality industry. *Contemporary Management Research*, *13*(2).

Torii, M., Fan, J. W., Yang, W. L., Lee, T., Wiley, M. T., Zisook, D. S., & Huang, Y. (2015). Risk factor detection for heart disease by applying text analytics in electronic medical records. *Journal of biomedical informatics*, *58*, S164-S170.

Tsai, C. H., & Brusilovsky, P. (2019). Explaining recommendations in an interactive hybrid social recommender. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 391-396).

Tussyadiah, I., & Zach, F. (2015). Hotels vs. peer-to-peer accommodation rentals: Text analytics of consumer reviews in Portland, Oregon. *Peer-to-Peer Accommodation Rentals: Text Analytics of Consumer Reviews in Portland, Oregon (April 10, 2015)*.

Vall, A., Dorfer, M., Eghbal-Zadeh, H., Schedl, M., Burjorjee, K., & Widmer, G. (2019). Feature-combination hybrid recommender systems for automated music playlist continuation. *User Modeling and User-Adapted Interaction*, *29*(2), 527-572.

Wang, H., Wang, N., & Yeung, D. Y. (2015, August). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1235-1244).

Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, *105*, 87-95.

Wei, C. H., Kao, H. Y., & Lu, Z. (2013). PubTator: a web-based text mining tool for assisting biocuration. *Nucleic acids research*, *41*(W1), W518-W522.

Wu, M.Y., Wall,G.,& Pearce,P.L.(2014). Shopping experiences: international tourist in Beijing's silk market. Tourism Management,41,96-106

Xia, L., Goldberg, D. M., Hong, S., & Garvey, P. (2019). Transfer Learning in Knowledge-Intensive Tasks: A Test in Healthcare Text Analytics.

Xiang, Z., & Krawczyk, M. (2016). What does hotel location mean for the online consumer? Text analytics using online reviews. In *Information and Communication Technologies in Tourism 2016* (pp. 383-395). Springer, Cham.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, *44*, 120-130.

Xiang, Z., Schwartz, Z., & Uysal, M. (2015). What types of hotels make their guests (un) happy? Text analytics of customer experiences in online reviews. In *Information and communication technologies in tourism 2015* (pp. 33-45). Springer, Cham.

Yuan, H., Lau, R. Y., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems*, *91*, 67-76.

Zainol, Z., Nohuddin, P. N., Mohd, T. A., & Zakaria, O. (2017, April). Text analytics of unstructured textual data: a study on military peacekeeping document using R text mining package. In *International Conference on Computing and Informatics* (pp. 1-7).

Zainol, Z., Jaymes, M. T., & Nohuddin, P. N. (2018, May). VisualUrText: A text analytics tool for unstructured textual data. In *Journal of Physics: Conference Series* (Vol. 1018, No. 1, p. 012011). IOP Publishing.

Zaki, M., & McColl-Kennedy, J. R. (2020). Text mining analysis roadmap (TMAR) for service research. *Journal of Services Marketing*.

Zenkert, J., & Fathi, M. (2016, May). Multidimensional knowledge representation of text analytics results in knowledge bases. In *2016 IEEE International Conference on Electro Information Technology (EIT)* (pp. 0541-0546). IEEE.

Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, *99*, 238-248.

Zheng, X., Zhu, S., & Lin, Z. (2013). Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach. *Decision Support Systems*, *56*, 211-222.

Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., & Yan, X. (2018). Measuring customer agility from online reviews using big data text analytics. *Journal of Management Information Systems*, *35*(2), 510-539.

Zhou, S., Ng, S. T., Lee, S. H., Xu, F. J., & Yang, Y. (2019). A domain knowledge incorporated text mining approach for capturing user needs on BIM applications. *Engineering, Construction and Architectural Management*.